Paradox Problems as a Tool for Understanding Statistical Reasoning

Andrew Neath* Bailey Farrar†

Abstract
Thought experiments in the form of paradox problems are useful for illustrating how our statistical reasoning may be correct, or how it may be flawed, and how our models for behavior in the sciences may be appropriate, or how such models may be incomplete. For example, the Monte Hall problem, the Exchange paradox, the Ellsberg paradox, and transposed conditional paradoxes are all interesting exercises in decision making, quantifying uncertainty, and statistical inference. In this paper, we use the Exchange paradox to illustrate what paradox problems can teach us about the best practices in statistical science.

Key Words: statistics education, statistical inference, decision science, behavioral science, Bayesian statistics

1. Introduction

Teaching proper notions of uncertainty is among the greatest challenges for statistical educators, consultants, and collaborators. This challenge is made difficult because important aspects of probability are not particularly intuitive. Persi Diaconis is credited with the observation “Our brains are just not wired to do probability problems very well.” If we are to successfully communicate notions of uncertainty in educating statistical scientists, we need a better understanding of how an over reliance on intuition may lead to flawed statistical reasoning.

Thought experiments have proven to be useful for exposing the flaws in our intuitive thinking. Kahneman (2011) provides an overview of the groundbreaking work in behavioral economics used to study heuristic biases in judgement and decision making. These experiments in social psychology have led to improved models of economic behavior. Perhaps similar types of thought experiments can be used to highlight where models for behavior in the statistical sciences need change.

As a working definition, consider a paradox problem to be a thought experiment which demonstrates how our statistical reasoning may go wrong. Perhaps the most famous paradox in statistics is the Monty Hall problem. Rosenhouse (2009) gives a book length treatment of what this problem can teach us about proper statistical reasoning. In this paper, our focus will be on the Exchange paradox, and the deeper lessons this thought experiment can provide. Section 2 provides the details on the Exchange paradox problem, and offers a Bayesian resolution. In Section 3, we take a deeper look at where our intuitive thinking failed us, what we can learn from our mistake, and how this lesson can improve our statistical thinking and communication. The paper closes with some final comments, including an argument on the importance of persistently examining the role that statistics plays in the scientific process.

*Southern Illinois University Edwardsville, Department of Mathematics and Statistics, Edwardsville, IL, aneath@siue.edu
†Southern Illinois University Edwardsville, Department of Mathematics and Statistics, Edwardsville, IL

2. The Exchange Paradox

In this thought experiment, you are to randomly select one of two sealed envelopes; knowing that one envelope contains twice as much cash as the other (or that one envelope contains half as much cash as the other). After opening the envelope, you are given the option to switch. If your envelope contains x dollars, then the other contains either $x/2$ or $2x$ dollars. It seems that from an expected value calculation

$$\begin{aligned} E &= \frac{x}{2}\left(\frac{1}{2}\right) + 2x\left(\frac{1}{2}\right) \\ &= \frac{5}{4}x > x, \end{aligned} \tag{1}$$

you should gladly accept an offer to switch envelopes.

Note that the calculation in (1) holds for any x, so you would not even need to open your envelope before agreeing to switch. But now the same line of reasoning holds after switching, so you should be agreeable to switching back. Under this logic, switching will continue ad infinitum. Clearly, something must have gone wrong with our reasoning in expression (1). What seems like a simple problem is showing itself to require a deeper level of thinking.

A Bayesian view of the Exchange paradox will solve the puzzle (Christensen and Utts, 1992). Let's start our more careful approach with some notation. Define θ as the smaller of the amounts placed in the envelopes. Let $\pi(\theta)$ denote a prior distribution on θ. We observe x, the amount in the selected envelope. Define A as the amount in the other envelope. Then the expected value calculation we need for investigating a switching strategy is

$$E(A|x) = \frac{x}{2}\,P\left(A = \frac{x}{2}\,|\,x\right) + 2x\,P(A = 2x\,|\,x). \tag{2}$$

It can be shown that $P(A = 2x\,|\,x)$ depends on the prior π over parameter θ as

$$P(A = 2x\,|\,x) = \frac{\pi(x)}{\pi(x) + \pi(x/2)}. \tag{3}$$

It makes sense that a decision to switch should depend on the amount observed from our envelope, and on our prior beliefs as to what dollar amounts are likely to be placed in the envelopes. For instance, if our prior is skewed toward larger amounts, and observed x is a relatively small amount, we are willing to switch. If our prior information is that smaller amounts seem more likely, we would tend to decide against switching envelopes.

3. What Can We Learn?

We now have an answer to a problem our intuition failed to solve correctly. Paradox problems like this traditionally have a place in mathematics and statistics as a fun diversion. However, let's take the problem more seriously and see what we can learn about where our intuition failed us. The derivation of the posterior probability in expression (3) follows from deriving the likelihood function in terms of unknown A as

$$L(A) = \begin{cases} 1/2 & \text{if } A = x/2 \\ 1/2 & \text{if } A = 2x. \end{cases} \tag{4}$$

We interpret a likelihood function as representing the data evidence. So, the data provides equal support to two possible values of A. It must be made clear that $L(A)$ does not represent a probability on A, nor does it represent any measure of belief.

Warnings on how a likelihood function does not represent a probability commonly accompany an introduction to likelihood inference. It is a tribute to the strong pull to think likelihood and probability are synonymous that this mistake is so commonly made. In fact, this misinterpretation of the likelihood function is the reason for misinterpreting the Exchange paradox problem. The expected value in (1) was incorrectly computed using $L(A)$ from (4), not $P(A|x)$ from (3). Our intuition led us treat evidence as equivalent to belief. This heuristic bias created a flaw in our statistical reasoning. As an aside, note from (3) that $P(A = x/2 \,|\, x) = P(A = 2x \,|\, x) = 1/2$ if and only if θ follows an improper uniform distribution. Because posterior probability is computed from both prior information and data evidence, it is reasonable that the case of no prior information is where posterior probability matches the data evidence. See Pawitan (2013) for a formal treatment of likelihood inference and its connection to Bayesian inference.

Our lesson from the Exchange paradox is how easy it is to mistake evidence for belief. This seems like a simple lesson, but failure to follow the correct line of statistical reasoning has enormous consequences. Misinterpretations of statistical significance, and p-values, in hypothesis testing problems are due in a large part to treating data evidence as equivalent to a probabilistic measure of belief. P-values, and hence determinations of statistical significance, are data evidence measures in the same category as likelihood inference (Neath, 2017). For statistically significant testing outcomes, the likelihood function is weighted more heavily toward one of the hypotheses, whereas the likelihood function in the Exchange paradox problem is weighted evenly. The error in treating evidence as belief is the same in both problems.

A claim of strong belief on the basis of a small p-value in a hypothesis testing setting is a consequential error in statistical reasoning, and a major contributor to the replication crisis in science (Ioannides, 2005; Open Science Collaboration, 2015). Concerns over the harm caused to science necessitated a need for the ASA to issue a statement clarifying the appropriate use of p-values in hypothesis testing (Wasserstein and Lazar, 2016). It seems that any insight as to how statistical reasoning goes wrong in this regard would be helpful in teaching good scientific principles. Indeed, the Exchange paradox can provide such insight.

4. Conclusion

In this paper, we focused on one aspect of one paradox problem. A review of the literature will reveal dozens of papers on the Exchange paradox alone, suggesting many different lessons in statistical reasoning. Our attention has been on what can be learned about p-values, their interpretations, and the their highly influential role in the scientific process. Correct statistical reasoning is the driving force behind doing good science, so it is important for statistical educators to carefully examine which scientific practices should be promoted, and which practices should be discouraged. If thought experiments aid us in this endeavor, then they can be as valuable in studying statistical behavior as they have been in studying economic behavior.

REFERENCES

Christensen, R., and Utts, J. (1992), "A Bayesian Resolution of the Exchange Paradox," *The American Statistician*, 46, 274-276.

Ioannides, J. (2005), "Why Most Published Research Findings are False," *PLoS Medicine*, 2,8,e124.

Kahneman, D. (2011), *Thinking, Fast and Slow*, New York: Farrar, Straus, and Giroux.

Neath, A. (2016), "P-values and the Likelihood Principle," *Proceedings of the American Statistical Association*, 966-971.

Open Science Collaboration (2015), "Estimating the Reproducibility of Psychological Science," *Science*, 349, aac4716.

Pawitan, Y. (2013), *In All Likelihood: Statistical Modelling and Inference Using Likelihood*, Oxford University Press.

Rosenhouse, J. (2009), *The Monty Hall Problem: The Remarkable Story of Math's Most Contentious Brainteaser*, Oxford University Press.

Wasserstein, R. and Lazar, N. (2016). "The ASA's Statement on P-values: Context, Process, and Purpose," *The American Statistician*, 70, 129-133.